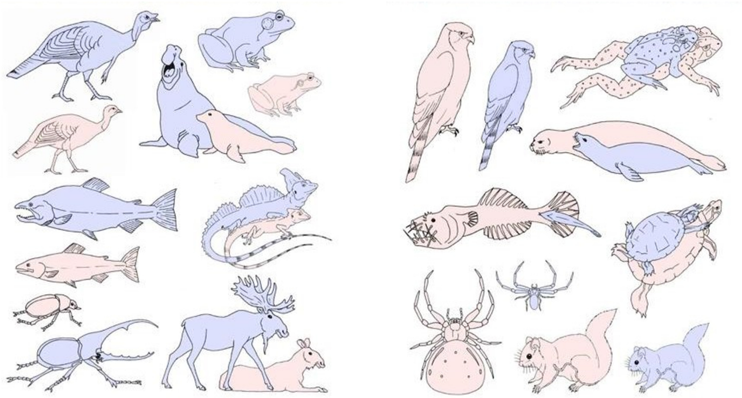
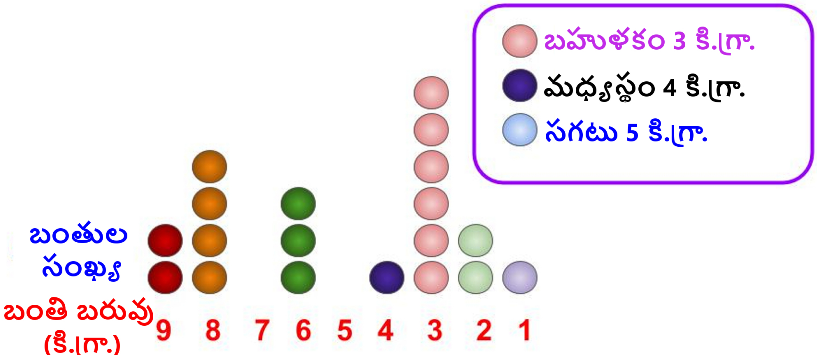
సగటు అంటే ఏమిటి?



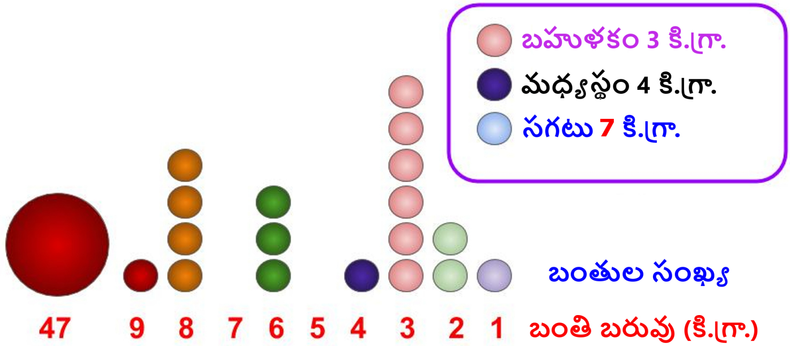
మానవజాతి లాగే జింకలూ, సింహాలూ మొదలైన అనేక ఇతర జంతువులలో మొగ జంతువు ఎత్తూ బరువూ ఆడ జంతువు ఎత్తూ బరువూ కంటే సాధారణంగా ఎక్కువ ఉంటాయి. అనేక ఇతర జంతువులలో ఇది తారుమారై ఆడ జంతువు ఎత్తూ బరువూ ఎక్కువ ఉంటాయి. కానీ అన్ని మొగ జింకలూ అన్ని ఆడ జింకల కంటే పెద్దగా ఉంటాయని అనలేం. ఉదాహరణకు పిల్ల జంతువులు చిన్నవే కదా. వయసు పెరిగిన జంతువులలో అన్నిటి ఎత్తూ బరువు ఒకటే ఉండదు. కొన్ని ఆడ జింకలు కొన్ని మొగ జింకలకంటే కొంత పెద్దవిగా ఉండవచ్చు. సగటున మొగ జింకలే పెద్దవి అనిపిస్తుంది. ఈ ఉదాహరణ సగటు అన్న మాటకి ఉజ్జాయింపు అర్థం ఇస్తుంది. కానీ ఇక్కడ కొన్ని అంటే ఎన్ని? కొంత అంటే ఎంత? ఇలాంటి ప్రశ్నలే ప్రతి ప్రయోగ ఫలితానికీ వర్తిస్తాయి. ప్రయోగశాలలో అదే కొలతని మళ్ళీ మళ్ళీ చేస్తే ఖచ్చితంగా అదే సంఖ్యా విలువ లభించదు. రెండు ప్రయోగాల ఫలితాలని పోల్చి చూడాలి అంటే వాటిని మళ్ళీ మళ్ళీ చేసినప్పుడు వచ్చే తేడాలని కూడా పోల్చి చూడాలి. అందువల్ల సగటుకి ఉజ్జాయింపు అని కాక ఒక గణిత రూపమైన నిర్వచనం చాలా అవసరం.

రెండు అంకెలని కూడి రెండుతో భాగిస్తే వచ్చే సంఖ్యని సగటు అంటామని అందరికీ తెలుసు. 2,4ల సగటు 3. ఈ పద్ధతిని ఎక్కువ అంకెలకీ ఉపయోగించవచ్చు. 2,4,6 ల యొక్క సగటు

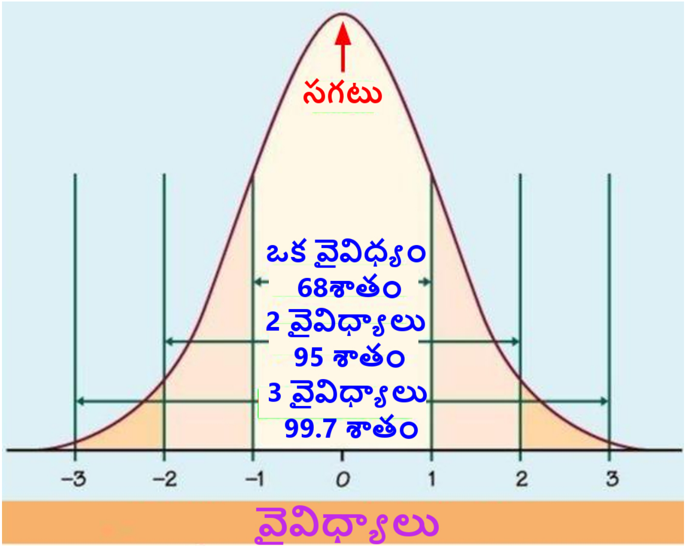


2+4+6=12÷3= 4. 1,5 లకి సగటు 3. ఏతావాతా సగటు విలువ నుంచి అసలు అంకెల గురించి ఏమీ తెలియదు. పై చిత్రంలో ఒక కిలోగ్రాము నుంచి 9 కిలోగ్రాముల వరకు బరువు ఉన్న 19 బంతులని చూడగలం. సగటు బంతి బరువు 5 కిలోగ్రాములు. కానీ ఉన్న 19 బంతులలో దేని బరువూ 5 కిలోగ్రాములు కాదు. 3 కిలోగ్రాముల బరువు ఉన్న బంతులు 6. ఆ బరువు ఉన్నవే ఎక్కువ. ఇలా బహుళంగా కనపడేదానికి బహుళకం ఇంగ్లీష్ లో మోడ్ అని పేరు. సగటు ఇంగ్లీష్ లో ఏవరేజ్ లెదా మీన్ కంటే బహుళకం ఉన్న బంతుల పంపిణీ గురించి కొంత సమాచారం ఇస్తోంది. బంతులని బరువు ప్రకారం ఆరోహణ క్రమంలో చూస్తే 4 కిలోగ్రాముల బరువు ఉన్న బంతికి ఒక విధంగా ప్రధానత కనపడుతుంది. తొమ్మిది బంతులు దానికన్నా ఎక్కువ బరువు వున్నాయి. అదేవిధంగా 9 బంతులు దానికన్నా తక్కువ బరువు ఉన్నాయి. ఇలా మధ్యలో ఉంది కనుక దీనిని మధ్యగతం లేదా మధ్యస్థం ఇంగ్లీష్ లో మీడియన్ అని పేరు.

ఉన్న బంతులలో కేవలం ఒక దానిని మారిస్తే సగటు విలువ మారిపోతుంది. ఉదాహరణకు ఒక 9 కిలోగ్రాము బరువు ఉన్న బంతి స్థానంలో 47 కిలోగ్రాముల బరువున్న దానిని ఉంచితే పక్క పేజీలో చిత్రంలో చూపినట్టు సగటు 5 కిలోగ్రాముల నుంచీ 7 కిలోగ్రాములకి పెరుగుతుంది కానీ బహుళకమూ మధ్యగతమూ మారవు. మిగతా వాటికంటే చాలా ఎక్కువ బరువు ఉన్న బంతి పంపిణీ లోకి వచ్చిందని తెలుసుకోలేము. పంపిణీని ఇంగ్లీష్ లో

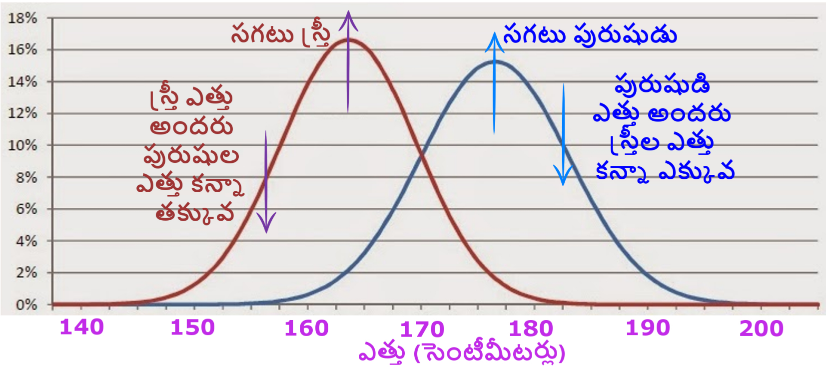


డిస్ట్రిబ్యూషన్ అంటారు. పంపిణీని పరీక్షించి తెలుసుకున్న బహుళక మధ్యగతాల కంటే సగటు గణించటం ఉపయోగకరం. సగటుకీ ప్రతి బంతికీ వ్యత్యాసాలు ఉంటాయి. ఈ సంఖ్యలతో వైవిధ్యం, స్టాన్డర్డ్ డీవియేషన్ అనే ఉపయోగకరమైన సంఖ్యని గణించవచ్చు. వైవిధ్యం వ్యత్యాసాల వర్గాల సగటుకి మూలం. వైవిధ్యం ఎక్కువ ఉంది అంటే ఉపయోగించిన పంపిణీలో ఏదో లోటు ఉందని తెలుస్తుంది. వ్యత్యాసం యొక్క వర్గం అంటే వ్యత్యాసాన్ని దానితోనే హెచ్చు వేయటం ముఖ్యం. అప్పుడు వ్యత్యాసం ఎక్కువ ఉన్న ఫలితాలు వల్ల వైవిధ్యం ఎక్కువగా పెరుగుతుంది. మొదట చూసిన 19 బంతుల పంపిణీకి సగటు 5 కిలోగ్రాములు, వైవిధ్యం 2.8. పెద్ద 39 కిలోగ్రాముల బంతిని పంపిణీ లోకి తీసుకున్నాక సగటు 7 కిలోగ్రాములూ వైవిధ్యం 8. సగటునీ వైవిధ్యంతో పోలిస్తే పంపిణీ లోటు పాట్లు తెలుస్తాయి. పంపిణీలో వ్యత్యాసాలు ఎక్కువ ఉన్నాయని చూసిన తరువాత, పంపిణీ నుంచి ఎక్కువ వ్యత్యాసం ఉన్న కొలతలని మాత్రం తొలిగించటం అవైజ్ఞానికం. పంపిణీని తయారు చేయటంతో ఉపయోగించిన ప్రయోగాలనీ పద్ధతులనీ మార్చడం వైజ్ఞానికం. ఆధునిక విజ్ఞానాన్ని సవాలు చేసేవాళ్ళు తమ సిద్ధాంతానికి ఎక్కువ వ్యత్యాసం ఉన్న కొలతలని తొలగించటానికి ఏదో ఒక హేతువు వెతుకుతారు. సమరూప గుంపులని పోల్చి చూడటం ఆధునిక వైజ్ఞానిక వైద్య పద్ధతిలో మాత్రమే జరుగుతుందని ఇంతకు ముందు చూశాము. దీనికి హేతువు ఎక్కువ వ్యత్యాసం ఉన్న కొలతలని తొలిగించరాదనే.



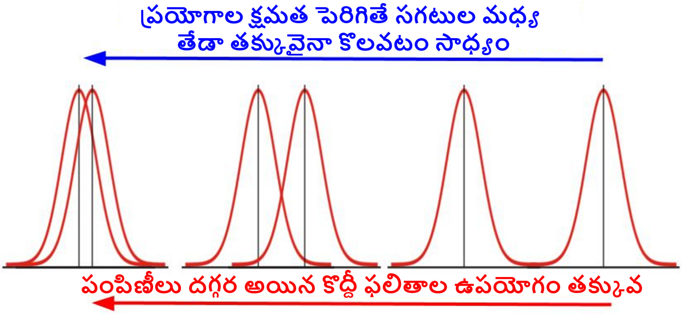
వైవిధ్యం యొక్క ఉపయోగం పంపిణీ గాసియన్ అయినప్పుడు కనపడుతుంది. జర్మన్ గణిత భౌతిక శాస్త్ర ప్రముఖుడు కార్ల్ గాస్ ఆవిష్కరించిన ఈ పంపిణీలో సగటు, బహుళక మధ్యగతాలు మూడూ సమంగా ఉంటాయి. అంతేకాక వ్యత్యాసం పెరిగిన కొద్దీ పంపిణీలో కొలతలు ఉండే సంభావన ఒక నిర్దిష్ట గణిత సూత్రాన్ని అనుసరించి తగ్గుతుంది. గాసియన్ పంపిణీలో 68 శాతం కొలతలు సగటుకి ఒక వైవిధ్యం ఎక్కువ తక్కువల మధ్య ఉంటాయి. అలాగే 95శాతం కొలతలు సగటుకి రెండు వైవిధ్యాల ఎక్కువ తక్కువల మధ్య ఉంటాయి. పైన చిత్రంలో చూపినట్లు ఘంట ఆకారంలో ఉండటం వల్ల దీనిని ఘంటాకార వక్రరేఖ, బెల్ కర్వ్ అని కూడా పిలుస్తారు.

నిజజీవితంలో గాసియన్ పంపిణీకి మంచి ఉదాహరణ ఈ అధ్యాయం మొదట్లో చర్చించిన స్త్రీ పురుషుల ఎత్తు కొలతలే. పురుషుల సగటు ఎత్తు 178.4 సెంటీమీటర్లు. వైవిధ్యం 7.6 సెంటీమీటర్లు. అదే స్త్రీల సగటు ఎత్తు 164.7 సెంటీమీటర్లు, వైవిధ్యం 7.1 సెంటీమీటర్లు. పక్క పేజీలో చిత్రంలో చూపినట్లు రెండు పంపిణీలూ చక్కటి గాసియన్ లు. స్త్రీల ఎత్తు పంపిణీ



వైవిధ్యం తక్కువ అవటంవల్ల ఆ గాసియన్ వెడల్పు కొంచెం తక్కువ ఉందని చూడగానే మనకి తెలిసి పోతుంది. 68 శాతం పురుషుల ఎత్తు 170.8, 186 సెంటీమీటర్ల మధ్య ఉంటుంది. అదే 68 శాతం స్త్రీల ఎత్తు 157.6 171.8 సెంటీమీటర్ల మధ్య ఉంటుంది. సగటు పురుషుల ఎత్తు ఎక్కువ. సగటున పురుషుల ఎత్తు ఎక్కువ అయ్యే సంభావన పంపిణీ పెద్దది అయిన కొద్దీ అంటే చేసిన కొలతల సంఖ్య పెరిగిన కొద్దీ 100 శాతానికి దగ్గర అవుతుంది. కానీ జనాభా లో ఒక అనియమిత జంటని అంటే ఒక స్త్రీనీ ఒక పురుషుడినీ ఏ పద్ధతీ నియమమూ లేకుండా ఎంచుకుంటే, ఏ విధమైన సందేహమూ లేకుండా ఆ పురుషుడి ఎత్తు ఆ స్త్రీ ఎత్తు కంటే ఎక్కువ ఉంటుందని చెప్పగలమా? ఎప్పటికీ అలా ఉండదు. చిత్రంలో చూపించినట్లు కేవలం కొద్ది మంది పురుషులే అందరు స్త్రీలకంటే పొడుగుగా ఉంటారు. అలాగే కేవలం కొద్ది మంది స్త్రీలే అందరు పురుషులకంటే పొట్టి.

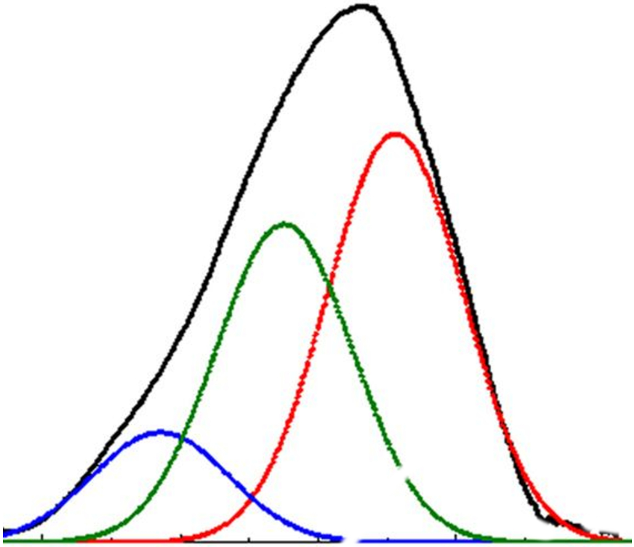
ఒక అనియమిత జంటలో స్త్రీ ఎత్తు ఎక్కువ అవటానికి కొంత సంభావన ఉంటుంది. దీనిని విపరీత సంభావన ఇంగ్లీష్ లో రివర్స్ పాస్సిబిలిటీ అని చెప్పవచ్చు. సగటు పురుషుడి ఎత్తుకీ సగటు స్త్రీ ఎత్తుకీ వ్యత్యాసం 13.7 సెంటీమీటర్లు. ఇది ఆ రెండు వైవిధ్యాల సమిష్టి విలువ 14.7 సెంటీమీటర్ల కంటే కొద్దిగా తక్కువ. అందువల్ల ఈ ఉదాహరణలో విపరీత సంభావన తక్కువగానే ఉంటుంది కానీ దాని విలువని గణించటానికి చాలా



క్లిష్టమైన గణిత సూత్రాలు అవసరం. స్థూలంగా రెండు సగటుల మధ్య వ్యత్యాసం తగ్గిన కొద్దీ విపరీత సంభావన పెరిగుతుంది. సగటుల మధ్య వ్యత్యాసం సూన్యం అయితే విపరీత సంభావన 50 శాతం అవుతుంది.

ఈ విపరీత సంభావనని గుర్తిస్తే సమాజంలో చాలా ఘర్షణలు తగ్గుతాయి. రెండు గాసియన్ ల సగటుల మధ్య వ్యత్యాసం ఎంత తక్కువైనా పంపిణీలని పెద్దవి చేయటం ద్వారా లేదా ప్రయోగాల క్షమతని పెంచటం ద్వారా దేని సగటు ఎక్కువో ఖచ్చితంగా చెప్పటం సాధ్యమౌతుంది. ఈ విషయాన్ని పైన ఉన్న చిత్రం స్పష్టం చేస్తుంది.

కానీ ఈ వైజ్ఞానిక సత్యం యొక్క ఉపయోగం ముఖ్యంగా సమాజ నియమాల విషయంలో తగ్గుతుంది. ఉదాహరణకు సగటున మొదటి గుంపు రెండవ గుంపు కంటే బుద్ధిలోగానీ, క్రమశిక్షణలో గానీ మెరుగు అని ఎవరైనా వైజ్ఞానిక సాక్ష్యం చూపితే, అది వివాదలకు మాత్రమే చోటు ఇస్తుంది. అదే సాక్ష్యం ఆధారంగా విపరీత సంభావన ఇంచుమించు 50 శాతం ఉంది కనుక దాని ఉపయోగం లేదని గుర్తించితే వివాదం లేదు. అలాగే ఇటువంటి వైజ్ఞానిక సాక్ష్యం ఆధారంగా పాఠశాలలనీ బందీఖానాలనీ మనం ఏర్పరిస్తే సరియైన ఫలితాలు రావు. విఫలమవటానికి విపరీత సంభావన ఇంచుమించు 50 శాతం కనుక. కానీ ఈ ఏర్పాట్లను అతిగా నమ్మి విఫలమవటానికి దోషులను వెతకటం వల్ల మళ్ళీ వివాదాలూ ఘర్షణలూ వస్తాయి.

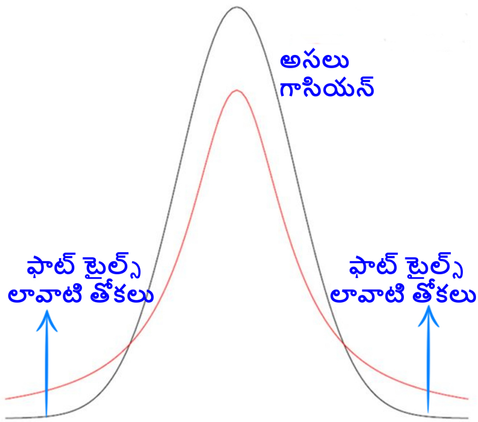


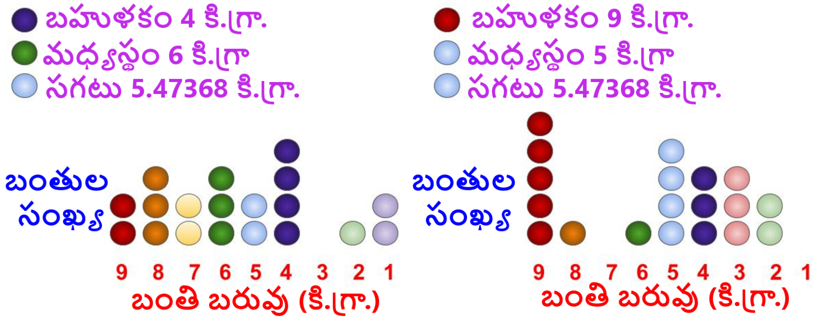
అసలు స్త్రీ పురుషుల ఎత్తు కొలతలు గాసియన్ రూపంలో ఎందుకు ఉన్నాయి? అన్ని పంపిణీలూ గాసియన్ లేనా? స్త్రీ పురుషుల ఎత్తు జన్యువులు, పోషణ, జబ్బులు మొదలైన అనేక విషయాల మీద ఆధారపడి ఉంటుంది. అలాగే ప్రయోగశాల ఫలితాలు, చేసే వ్యక్తి యొక్క సామర్థ్యం, చేస్తున్న సమయంలో వాతావరణం, యంత్రాల వయస్సు, వాటిని సంరక్షణ వగైరా అనేక స్వతంత్ర కారణాల వల్ల మారవచ్చు. ఇవన్నీ ఒక్కోకటీ గాసియన్ పంపిణీలు కాకపోయినా అనేక పంపిణీలు కలిస్తే వాటి సమిష్టి పంపిణీ మాత్రం గాసియన్ అవుతుందని గణిత సూత్రల ఆధారంగా నిరూపించవచ్చు.

ప్రాథమిక రసాయన భౌతిక శాస్త్రాలకి గణిత సంబంధాలు ఉన్నాయని ఇంతకుముందు చూశాం. అందువల్ల ప్రయోగ ఫలితాలు గాసియన్ ఆకారంలో లేకపోతే ఆ ఫలితాలని కొన్ని గాసియన్ ల సమిష్టిగా గణిత సహాయంతో విడగొట్టటం సాధ్యం. దీని పై చిత్రంలో చూడవచ్చు. కానీ ఆధునిక వైద్య పరిశోధనలో, పర్యావరణ విషయంలో సైతం ఇలాంటి సులువు లేదు. ఇక సమాజ విజ్ఞాన శాస్త్రం లో అసలు ప్రాథమిక గణిత సూత్రాలు

లేవు. ఇంతకుముందు కూడా చర్చించిన ఒక విషయం చాలా ముఖ్యం. భౌతిక రసాయన ప్రయోగాలు పూర్తిగా సమరూపత ఉన్న కోట్ల కోట్ల బణువుల మీద చేయబడతాయి. అందువల్ల ప్రయోగ పరికరాలని మెరుగుపరచగలిగితే పంపిణీ వైవిధ్యం తగ్గించవచ్చు. వైవిధ్యం బణువుల లక్షణం కాదు. ఆ ప్రయోగ పరికరాల పరిమితి. వైద్య పర్యావరణ సమాజ శాస్త్ర విషయంలో వైవిధ్యం పరికరాల పరిమితి కానే కాదు. విషయంలో క్లిష్టత, సమరూపత ఉన్న గుంపులని ఎంచుకోవటంలో ఇబ్బందుల కారణంగా వైవిధ్యం ఎక్కువ ఉంటుంది. కేవలం పంపిణీ పెద్దది చేయటం వల్లా పరికరాలని మార్చటం వల్ల చాలా విషయాలలో వైవిధ్యం తగ్గదు.

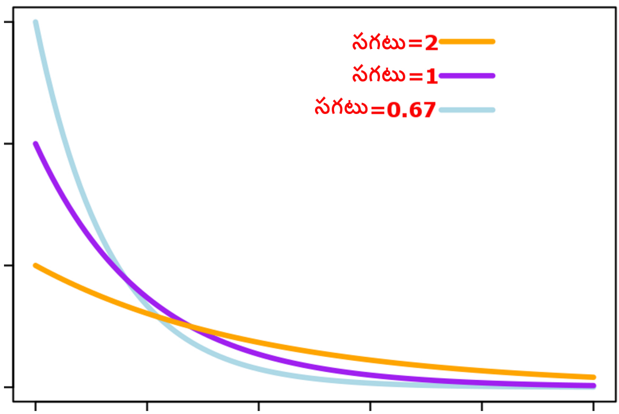
గాసియన్ పంపిణీలో సగటుకి మూడు వైవిధ్యాల కన్నా ఎక్కువ వ్యత్యాసం ఉండటానికి సంభావన చాలా తక్కువ. ఉదాహరణకు ఇంతకుముందు ఉదాహరించిన సగటు, వైవిధ్యం సంఖ్యల ఆధారంగా వెయ్యి మంది పురుషులలో ఒక్కడు మాత్రమే రెండు వందల సెంటీమీటర్ల కంటే ఎత్తు ఉంటాడని చెప్పవచ్చు. నిజ జీవితంలో చాలా పంపిణీలు స్థూలంగా చూస్తే గాసియన్ లాగా కనపడతాయి కానీ సగటుకి చాలా ఎక్కువ వ్యత్యాసంలో సంభావన గాసియన్ కంటే చాలా ఎక్కువ ఉంటుంది. వీటిని ఫాట్ టైల్ డిస్ట్రిబ్యూషన్, లావాటి తోకలు ఉన్న పంపిణీలు అని





అంటారు. ముఖ్యంగా ఆర్థిక రంగంలో ఈ పంపిణీలు చాలా అలజడి కలిగిస్తాయి. సగటుకి అంత వ్యత్యాసం ఉంది కనుక ఆ ప్రమాదం జరిగే సంభావన లేదు అని వైజ్ఞానికులూ రాజయ నాయకులూ ధీమాగా ఉంటారు. పంపిణీ గాసియన్ కాక లావు తోకలతో ఉంటే ప్రమాదం జరిగి అందరికీ నష్టం కలుగుతుంది.

అధ్యయనం చేస్తున్న విషయం క్లిష్టత వలన ఫలితాల పంపిణీకీ గాసియన్ కి చాలా తేడా ఉండవచ్చు. రెండు పంపిణీలని పోల్చి చూడటమే కష్టం కావచ్చు. బంతులతో కృత్రిమంగా తయారు చేసిన రెండు పంపిణీలని చిత్రంలో చూడవచ్చు. రెండిటి సగటూ సరిగ్గా ఒకటే. వైవిధ్యంలో సైతం పెద్ద తేడాలేదు. బహుళక మధ్యగతాలలో తేడాలు ఉన్నాయి. కానీ వాటి ఆధారంగా సైతం ఈ రెండిటినీ పోల్చి ఒక దాన్ని ఎంచుకోవడం



అసాధ్యం. పంపిణీలు పూర్తిగా చూపించక లేదా వాటిని నమోదు చెయ్యకుండా మధ్యగతం ఎక్కువ ఉన్న పంపిణీని ఉత్తమమని తేల్చటం సబబేనా? పంపిణీలని పోల్చటంలో ఉన్న చికాకులు చాలనట్టు కొన్ని అతి సరళ పంపిణీలని పోల్చటం కష్టం కాదు, అసంభవం. పక్క పేజీలో కింది చిత్రంలోని ఎక్స్పొనెంషియల్ డిస్ట్రిబ్యూషన్ మంచి ఉదాహరణ. సగటు, వైవిధ్యం గణించటం కష్టం కానేకాదు. పెద్ద క్లిష్టతా కనపడదు. కానీ పోల్చటం? ఇటువంటి పంపిణీలు భౌతిక రసాయన శాస్త్రాల్లోనే కానీ మన సామాజిక జీవితంలో కనపడకపోవటం ఒక పెద్ద అదృష్టం.